

Nucleotide sequence of the *S. clavuligerus* chromosome including and flanking *cas1* (SEQ ID NO:1)

NcoI . . . . .

1

GGTACCGCCCGCCGCGACGGGGCCTCGGAGCCGGCCTGGCCACTGGTCCTGGTGGGGCC 60  
M A P P P Q G P A E A P G T V L V V G

61

ACCCTATCACCGGGCGGTGGGCGCGTCGTCGAGGGCCTGTGCCTGGGCACCCACACGC 120  
T P Y H G A V R R L L S G S V S G H T H

<orfup3 . . . . .

121

GCCTTTCCGGGCTCCGGCCCAGTGTCGGTGCCCATTCGCGCCACAGGAACGGGCGCAT 180  
\* L W P Y R A T D K G A

Y

A S L G P P R T M . . . . .

181

TAGCCCCAGGTCTATCTGCTTCCGGGCCACCTGCTCCTTCAGGGCGTGAGCATCTGGCA 240  
D P D L Y V F A R H V L F D R V E Y V

T

241

CGTGGTCGCGGGCCGCGGGTGAGCCCCAGTGGGCGGGCGGTGCCGGGCAGGGCCACGAG 300  
C W R G A A W E P D G A R W P G D R H

E

301

TGGCACCCACCGGGAGGCGCGCTCCTCAAGCCAGGGCCAGTCTTAGGTCAACTGCCT 360  
G H T T G E A A L L E T G T L I W N V

S

361

GGTGTCTACCACTAGCTCGCCTACCACGGGGGCTCCAGCAGCTTCTCGGCCCCGCTA 420

FIG. 1A

W L H H T I S R I T G G L D D F L R A  
I

421  
GAGCCTGAACGGGGCCCCGGTCTGGGGTGAACCCCTTCTTCTTCTGGCGCAGGAGCCGCTT 480  
E S K G R A L G W K P F F F V A D E A  
F

481  
CATCAGCTAGCGCCCCACGGCAGCGACGGCTGCGGCGGCAACAGCTTGCGGAACTTCAT 540  
Y D I A P T G D S G V G G N D F A K F  
Y

<orfup2  
541  
GCGCCACTACTGGCGGAACGCGACGAGCAGGCAGTATGGCCGGCTACGGTGCCTGTACTT 600  
A T I V A K R Q E D T M G A S A V S M

601  
TGCTGGAGGTCTCTAAGGCCACCGACACGACCCCGACGCCTTCCCCACAGGGGGCGCTT 660

661  
CCTGCCGCTGCGGCGCCTGCGGCGCCGGCAGAGGGGCCGCTGCCAGGGTCGCAGGAC 720

721  
CTCTCCCGAACC GCCCGCCGAAC TCGGGCACGACAGGGCGCCGAACGCCTT GCGCTTCATG 780

781  
GCCGGTCGCATGCCCCGAACGTGGCCTGCACATGCGGCCAGCCCTGGGGAGCATGGGGGC 840

841  
CTCGGCCGGCTGGGGCCGCCGAGGCCCCCATGCCTGCGCGGCCTGGCCGGGCTCGCTCGG 900

901  
CCTGCCCAGCCTGCCACGCGCACCAAGGCCACACAGCCTGTCGAGCCTGCCTGGCCTGCC 960

FIG. 1B

961  
ACGCGCACCAAGGCCACACAGCCTGTCGAGCCTGCCCAGCCTGCCACGCGCACCAAGGCC 1020

1021  
GTGCGGCCTGCCCAGTCAACGGCTAGTACCGCTCGTTACGGCCCCACATGGCGAGGGGCC 1080  
\* N G I M A L L A P T Y R E G

1081  
TGTGGCCCCACCCTCTAGCGCCGGCAGTGGAGGCGCTCCCTGGCCAGCAGGTCGGCCTAGC 1140  
S V P H S I A A T V E A L S R D D L R I

1141  
TCCGCCCGCGCTCTAACAGGCGCTCTACCCGGCCCAAGCGCCACGGGCCCTAGCCCTGCT 1200  
S A A A L N D A L H A P N A T G P I P V

1201  
GCAGGAGCGGGGCCACCACGTCGGTCCGCTCGCGCTCGACACGGTCCCAGTCGGGGTCTG 1260  
V D E G R H H L W A L A L Q A L T L G L

1261  
GCAGGCGCTGCCCCGCGTCGGCCACGTCGTTGCTCGCCAACGCGCGCTCCCGGCCCTCGCG 1320  
G D A V P R L R H L L S R N R A L A P A

1321  
ACTTGCCCCCGACCGGGGCCGCTTCAGGAGCAGGGGGTCTAGCAGCCACCACGCTTACC 1380  
S F R P Q G R R F D E D G L D D T T R I

1381  
ACGGCCACTCTTTTGGGGCAGGGTCTCCCCGCATTGCTGCTAGGGCTAGGGGTGAGGG 1440  
T G T L F G R G L P A Y A V I G I G L E

1441  
CCGTCTGCCCCGTGGTGGAGCAGGAGCTAGGGCGCGCTGGTGTCCGAGGTGAGCGAGACGT 1500  
R C V P V V E D E I G R S W L S W E S Q

FIG. 1C

1501

GGCGGCAGTGGCCCCACGTGGCGCAGGCGGGCCGCGTCGCACCGGCGCCTCCCGAGCCTCT 1560  
V A T V P H V A D A R R L T A A S P E S

1561

CTGGCTCGGACGCCTGGAACGGGAGCGCGTGGTCGAGCCGGTGGCGTGGGTGCCAGAGGA 1620  
L G L R R V K G E R V L E A V A G V T E

1621

GCTAGCCGTGGCGGGCCAGGCAGGTCACGACCATCATGTCCAGCTACGCCAGCCACGGCT 1680  
E I P V A P D T W H Q Y Y L D I R D T G

1681

CTGCTGCGTCCCTGGCAAGCGTCCGGCGCGCCTGCATCCTGCCGAGCGGCGTGTTCGGGA 1740  
L R R L S R E C A A R V Y S P E G C L G

1741

CCCTCCGCGGCAGCCTGCTCGCGTGGTACGGCTTGAACCACCGCTAGTCGTGGAGCAGGG 1800  
Q S A G D S S R V M G F K T A I L V E D

1801

CCGCCGGGCGCTGGCGGGCAGGCTCGTTCGAGGAGTGGCCGCGGCTCGGGGACCTGCAGCC 1860  
R R G A V A R G L L E E G A G L G Q V D

1861

GCCACAGGTCGTCCCACTGGGGCCGCAGCTGCCGCCGCGCCTACCACCGGCAGCGGGCCCC 1920  
A T D L L T V G A D V A A R I T A T A R

1921

GCGCCAGGCCCCGAGGCATCTTCAGCCACCAGCCGTCGTCGGCTCGGGGACCCGTGACT 1980  
A R D P R G Y F D T T P L C G L G Q A S

1981

GGCCTTCCAGGGCGTCCCGCGCCTGGCCGCGCTGCGCCTTGGCGCCGCGCCTGTGCCTTGGCC 2040  
V P L D R L A R V P P R P V A A S V S G

FIG. 1D

<orfup1

2041  
CCGGGGACTCGGGCGGAGAGCGGGACATACGGAACCTCCACAGGCGGAGCCGGGAACGGG 2100  
GGCCCCCTGAGCCCCGCTCTCGCCCTGTATGCCTTGGAGGTGTCCGCCTCGGCCCTTGCCC  
A P S E P P S R S M

2101  
ACGAGGGCGAGGACGGGACGGAACGAAGGAGAGGACGGGACGGACAGCACGGACGGGACG 2160  
TGCTCCCCTCCTGCCCTGCCTTGCTTCCTCTCCTGCCCTGCCTGTCGTGCCTGCCCTGC

2161  
GACGGAACGGAGTCGGGAACCGGGGGGGGTGACCGGAACCGGGCCGTCCTTGGCCCTCCC 2220  
CTGCCTTGCTCAGCCCTTGCCCCCCCCCACTGGCCTTGCCCCGGCAGGAACCGGGAGGG

2221  
CCGTCCTCCCCGCCATCCGCCGTTCCTCCCCCGTTCCTCTCCCGTCCTCCAGCCAACACC 2280  
GGCAGGAGGGGCGGTAGGCGGCAAGAGGGGGCAAGGGAGAGGGCAGGAGGTCGGTTGTGG

2281  
GCCGCCCTTTCCAAGCGCTTGACACGGCACCGACAGCCGCCGCCGGGCGCCCGATGGGGA 2340  
CGGCGGGAAAGGTTTCGCGAACTGTGCCGTGGCTGTCGGCGGGCGGCCGCGGGCTACCCCT

2341  
CCCGTGCCCCCGGTTGAGCGGCGGTGAGCGCCGGTACGGGACCCACGCGCCGCCGCCCCG 2400  
GGGCACGGGCGGCCACTCGCCGCCACTCGCGGCCATGCCCTGGGGTGCGCGGGCGGGCGGC

2401  
GGCGCCCGCCAGGGCCCGCGCGGCCACCCCGGCCCGCCCCGGCCGGAGCGGCGATCCGGG 2460  
CCGCGGGCGGTCCCCGGGCGCGCCGGTGGGGCCGGGCGGGGCGGCCTCGCCGCTAGGCCC

FIG. 1E

2461  
CCGCTCGCTGCAAGAGGAACATCCACAGCCGCACAAGGAGCGCTCCGCACAGTGGGCACC 2520  
GGCGAGCGACGTTCTCCTTGTAGGTGTCGGCGTGTTCCTCGCGAGGCGTGTCAACCGTGG

2521  
ACGTCCGCCCCGTCCCCACACCGTGGCCGGTCCCCACGGACAGCACAGCACCGCACAG 2580  
TGCAGGCGGGGCAGGGGGTGTGGCACCGGCCAGGGGTGGCCTGTCGTGTCGTGGCGTGTG

2581  
CACCACATCGCACGGCACAGCACAGCACCGGCCACGAGGAACCAAGGAAAGGAACCAC 2640  
GTGGTGTAGCGTGCCGTGTCGTGTCGTGGTGGCCGTGCTCCTTGGTTCCTTTCTTGGTG

cas1>  
M T S V D C T A Y G P E L R A L A A  
2641  
ACCACCATGACCTCAGTGGACTGCACCGCGTACGGCCCCGAGCTGCGCGCGCTCGCCGCC 2700  
TGGTGGTACTGGAGTCACCTGACGTGGCGCATGCCGGGGCTCGACGCGCGCGAGCGGGCGG

2701  
CGGCTGCCCCGGACCCCCGGGCCGACCTGTACGCCTTCCTGGACGCCGCGCACACAGCC 2760  
R L P R T P R A D L Y A F L D A A H T A

2761  
GCCGCTCGCTCCCCGGCGCCCTCGCCACCGCGCTGGACACCTTCAACGCCGAGGGCAGC 2820  
A A S L P G A L A T A L D T F N A E G S

2821  
GAGGACGGCCATCTGCTGCTGCGCGGCCCTCCCGGTGGAGGCCGACCCGACCTCCCCACC 2880  
E D G H L L L R G L P V E A D A D L P T

NcoI .

2881  
ACCCCGAGCAGCACCCCGGCGCCCCGAGGACCGCTCCCTGCTGACCATGGAGGCCATGCTC 2940

FIG. 1F

T P S S T P A P E D R S L L T M E A M L

KpnI.

2941

GGACTGGTGGGCCCGCGGCTCGGTCTGCACACGGGGTACCGGGAGCTGCGCTCGGGCACG 3000  
G L V G R R L G L H T G Y R E L R S G T

3001

GTCTACCACGACGTGTACCCGTCGCCCCGGCGCGCACCACCTGTCCTCGGAGACCTCCGAG 3060  
V Y H D V Y P S P G A H H L S S E T S E

3061

ACGCTGCTGGAGTTCCACACGGAGATGGCCTACCACCGGCTCCAGCCGAACCTACGTCATG 3120  
T L L E F H T E M A Y H R L Q P N Y V M

3121

CTGGCCTGCTCCCCGGGCGGACACGAGCGCACGGCGGCCACACTCGTCGCCTCGGTCCGC 3180  
L A C S R A D H E R T A A T L V A S V R

3181

AAGGCGCTGCCCCCTGCTGGACGAGAGGACCCGGGCGCGGCTCCTCGACCGGAGGATGCCC 3240  
K A L P L L D E R T R A R L L D R R M P

3241

TGCTGCGTGGATGTGGCCTTCCGCGGCGGGGTGGACGACCCGGGCGCCATCGCCCAGGTC 3300  
C C V D V A F R G G V D D P G A I A Q V

3301

AAACCGCTCTACGGGGACGCGGACGATCCCTTCCTCGGGTACGACCGCGAGCTGCTGGCG 3360  
K P L Y G D A D D P F L G Y D R E L L A

3361

CCGGAGGACCCCGCGGACAAGGAGGCCGTCGCCGCCCTGTCCAAGGCGCTCGACGAGGTC 3420  
P E D P A D K E A V A A L S K A L D E V

FIG. 1G

3421  
ACGGAGGCGGTGTATCTGGAGCCCGGCGATCTGCTGATCGTCGACAACTTCCGCACCACG 3480  
T E A V Y L E P G D L L I V D N F R T T

3481  
CACGCGCGGACGCCGTTCTCGCCCCGCTGGGACGGAAGGACCGCTGGCTGCACCGCGTC 3540  
H A R T P F S P R W D G K D R W L H R V

3541  
TACATCCGCACCGACCGCAATGGACAGCTCTCCGGCGGGCGAGCGCGGGGCGACGTCGTC 3600  
Y I R T D R N G Q L S G G E R A G D V V  
A F T P R G \* *SacI*

3601  
GCCTTCACACCGCGCGGCTGAGCTCCCGGGTCCGACACCGCGCGGCTGAACCCACGGTCC 3660  
CGGAAGTGTGGCGCGCCGACTCGAGGGCCCAGGCTGTGGCGCGCCGACTTGGGTGCCAGG

3661  
GGGGCCCACGGTCCGGCACCGCGCGGCTGAGCCCCCGGGTCCGGCAGCGGGCGGC'TGAAC 3720  
CCCCGGGTGCCAGGCCGTGGCGCGCCGACTCGGGGGCCCAGGCCGTCGCCCCGCCGACTTG

3721  
CCCCGCCCCGGGCCACCGCCCGACCGCCCCCGCGCACCGGACGCGCCCGCCTGTACGGCG 3780  
GGGGCGGGGCCCCGGTGGCGGGGCTGGCGGGGGCGCGTGGCCTGCGCGGGCGGACATGCCGC

3781  
GTCCCCGCCCGGGCCCCGTACACCTGAAGCGCCCCGGCGGACCGCCCGCCCCGCCGGGGGACGG 3840  
CAGGGCGGGCCCCGGGCATGTGGACTTCGCGGGCCGCTGGCGGGGGGCGGCCCCCTGCC

-----> <-----

3841  
ACAGAGCCGGGTGCGGGAGGACGTCCTCCCGCACCCGGCTCCACCGTTCCGCACCGACC 3900  
TGTCTCGGCCCACGCCCTCCTGCAGGAGGGCGTGGGCCGAGGGTGGCAAGGCGTGGCTGG

FIG. 1H

3901  
GCACCCGACCGTGCCCGCAGGCGCCACCGGCACCGCACCGCCCGCGCCGGCAGCCACCACA 3960

CGTGGGCTGGCACGGCGTCCGCGGTGGCCGTGGCGTGGCGGGCGCGGCCGTCGGTGGTGT

3961  
GGCGCCACGCCCGCCGCACGGTGCCCGCGCTGCTCAGCCCCCGTCCACCGGGCTGTCCAG 4020

CCGCGGTGCGGCGGGCGTGCCACGGGCGCGACGAGTCGGGGGCAGGTGGCCCCGACAGGTC

\* G G D V P S D

L

4021  
GTCGGCGGCGTCGCGCGGGGGCTACTTGAGGGCCAGCCCGCGGTGGGGGGCCTGGGGCG 4080

L R R L A G G I F E R D A A S G G S G

A

4081  
CTCTACGGGGGTGTGAGGGCCCTAGTGGAGGTCGCTCCGTATGCCGTCGTCTAGCCGGTG 4140

L H G W V G P I V E L S A Y P L L D A

V

4141  
GGCGAAGAGCAGGAGCTGCCGCTTTGTGTGCAGGTCCC CGGGCCGTCGTGGTGCCGGGC 4200

R K E D E V A F C V D L A G P L V V A

R

4201  
GCGGCACTGCCTCCGGTCGCGGCGGAGCTGCGAGGGGGCGGGGCCCACAGCGGGGGTG 4260

A T V S A L A A E V S G G A G P T A G

V

NcoI

4261  
TAGGCACAAGAGGGTCCACGCGTGGTACCACTCGTCTAGGCGCCGCGGGCCCGGGCCTCTC 4320

D T N E W T R V M T L L D A A G P G S

L

FIG. 1I

4321  
CTTCTGGACGAGGGTCTTCGGCCACTCCATGAGGAGCGCCACCGCTTTGGGTTCGAGGGC 4380  
F V Q E W F G T L Y E E R T A F G L E  
R

4381  
CACCCGTGCCGCGGGTCTTCCTTGCGCTCCAGGGGTGGGCGCTTGTGGGCGGGCG 4440  
H A R R A W F S R S T G W G A F V R G  
A

4441  
GCGGAAGCGGGGGCGAGGGGCGCAGCCGCGACTCGCGGCGCCGGTCTGGCCTGTCGTC 4500  
A K R G R E G A D A S L A A A L G S L  
L

4501  
CTGGTCCGACACGCCCCGACGAGTGGCCGCGGGGCGTCTAGCCCCGCTAGGCCGCGTGGTA 4560  
V L S H P S S V P A G C I P A I R R V  
M

4561  
GGGGCCTACGCTGTGCCGGGTGACCATCCGCACCCGGCGCGGGTAGCTGGTCGGGCACTG 4620  
G P H S V A W Q Y A H A A G M S W G T  
V

4621  
GTCCCGGTCAAGGGCATGGGGGTCGAGGAGCCACTCGTCGGCCACGACGCGGCGCTGTAA 4680  
L A L E R V G L E E T L L R H Q A A V  
N

4681  
CAGGACGCCTCACTAGTCGCCTTTTCGCCCTGGGGCTGCCCACCAACGGCCCGCTCGACCT 4740  
D Q P T I L P F R S G S P H N G P S S  
S

FIG. 1J

4741  
CTGGGGCAACGGCTTCTCAGGCCGCCACTGCTGCGTCATGGCGGGCCACAGGTCGCCGTC 4800  
V G N G F L G A T V V C Y R R T D L P  
L

4801  
GGGGCGTGGCTAGTCGGTCAGCATGGGCCACACCAGGGCCGGCTTCTTGCTGCCTGTCTC 4860  
G A G I L W D Y G T H D R G F F S P C  
L

4861  
GTGGTGCAAGCAGGGCAGCCGCAAGCCGCACGGCATGTACCGCATTGGCTAGGCCCGCAG 4920  
V V N T G D A N P T G Y M A Y G I R A  
D

<orfdwn1  
4921  
GGCGTCCTGGAGGGGCAGGTCGTTGCCGTCAAGCAGCTAGAGCTTATACGCCGTAAGGTG 4980  
R L V E G D L L P L E D I E F I R C E  
M

4981  
GCGACTGGAGGAACAAGCTAGGGGGGCTGTTGTCCAGCCAGCACCGGCCTCTGAGTCTC 5040  
\*

5041  
GGTCAACCCCCGCTAGAGCCACCGGGTGTCGAGGTCCGACGCGTCGACCTGTAGCACGCC 5100  
W N P A I E T A W L E L S R L Q V D H  
P

5101  
CTAGTCGGGCCCTCATGACCGTGACCTCGTCTATGAGGCCTAGCACGGCGAGGTGGTCGAA 5160  
I L G S Y Q C Q L L Y E P D H R E V L  
K

5161  
GAGCTAGTACGCCAACTACAGCAGGCCCCACGGCTGGGTGAGGTCGGGGGCCAGCTGGTC 5220

FIG. 1K

E I M R N I D D P T G V W E L G R D V  
L

5221

CCAGAACATCAGGCTCGGCTAGCCTGGGCAGAGCGGCCAGCGCGCTCGCGGAGCCACTT 5280

T K Y D S G I P G T E G T A R L A E T

F

NcoI

5281

CGGGTACCCCGGCTTGGTCAAGAGCTTCTACTTCGGCGGCGGCCCTGCGGGTCACCAC 5340

G M P G F W N E F I F G G G R S A W H

H

5341

CCGGAGCGGCCTCAGGGCCCTCTGGTCCTGCAGGAAGTAGTGGGGCTGGGCGAGCGGGGC 5400

A E G S D R S V L V D K M V G V R E G

R

5401

GGCGTCCACGGCACCAGGGCGGCGGAGCCGGAGGAGGGCCATCTACAGGTAGTCGGCCCG 5460

R L T G H G A A E A E E R Y I D M L R

A

5461

CTGCTAGACCAGCAGCCACAAGTAGTCCTAGCCGTGGTGCGGGAGGGCCCGTGTCTTGGC 5520

V I Q D D T N M L I P V V G E R A C F

R

5521

CTTGACAGGAGTGACTTCGACTTGCCGACCTTCTGCCCGCCCCACCCCGCGACCATCCC 5580

F T D E S F S F P Q F V P P H P R Q Y

P

5581

GAACCCGCGCTACGGGTGGAGCGCCTACTGCGGCAAGAGCAGCTCCGGGGCCGGCATCGC 5640

K P A I G V E R I V G N E D L G R G Y

R

FIG. 1L

5641  
CGCGTGGCGGAGCATCCCCTTGAGGTCCAGGCCGTGGCCCTAGCAGGTGACGAGGGGCCT 5700  
R V A E Y P F E L D P V P I T W Q E G  
S

5701  
CACCCACTTGCAGAGCCAGCAGGTGCGGAAGAACTACTAGAGGGTCACGAGGAGCTTCTC 5760  
H T F T E T T W A K K I I E W H E E F  
L

5761  
CCGTGCTAACGCGGCCAGGGCGAGGGGCCGAGCCTGTCCACGGCGGCTGGGGCATGTG 5820  
A R N R R D R E G A D S L T G G V G Y  
V

5821  
GACGGGGTACTACAGCCGGGTCGCGAAGACCTTGGGCGCGCGCTAGGGCTGCTTCCGCGC 5880  
Q G M I D A W R K Q F G R A I G V F A  
R

5881  
CGGGGCCCCAGTACACCAGCTCGTAGCGGTCTAGGAGCCGGTCGGCGTCGCCTAACACGTC 5940  
G R T M H D L M A L D E A L R L P N H  
L

5941  
GCCGTCCTGCAACCGGTAGACCGGCTGGGCCTACACGGCCCAGACGTACGGCTCCATCTC 6000  
P L V N A M Q G V R I H R T Q M G L Y  
L

6001  
GGGGTCGTACTAGCCCAACAACCTCTGGAGCTTTGGGAGCCACACCTTCACCACGAGCCA 6060  
G L M I P N N S V E F G E T H F H H E  
T

FIG. 1M

6061  
CTTCCTGTCAGGGGTCATCGGCTCAAGCAGCCGGCGGACGCGGACGGCCCACTCGACGGC 6120  
F S L G W Y G L E D A A Q A Q R T L Q  
R  
. . . . .  
. . . . .  
6121  
CTCGTACAAGACCATCAAGACGCCTAACTGGGGGCGGTATGGGGCGACCTGGACGCGTAC 6180  
L M N Q Y N Q P N V G A M G R Q V Q A  
H  
. . . . .  
. . . . .  
6181  
ACTGCCGACCGTTGGCAGATAGAAGAGAATGGACTTCACCCTGGCTCTCCGGTTCGCGG 6240  
TGACGGCTGGCAACCGTCTATCTTCTCTTACCTGAAGTGGGACCGAGGAGGCCAAGCGCC  
S G V T P L Y F L I S K M  
. . . . .  
. . . . .  
6241  
CGCCCTCCATTGACGTGCGCCGAAAGCGGCTCGACCGTCCCACTCCGCCCTTGAGTTCCG 6300  
GCGGGAGGTAAGTGCACGCGGCTTTGCGCCGAGCTGGCAGGGTGAGGCGGGAACCTCAAGGC  
. . . . .  
. . . . .  
6301  
TCTGACGCCGCGCCAGTCGGCGGGCCGTCCGCCGGGGTGCCCGCGGGGTCCGCACCCGC 6360  
AGACTGCGGCGCGGTCAGCCGCCCGGCAGGCGGGCCCCACGGGCGGGCCCCAGGCGTGGGCG  
. . . . .  
. . . . .  
6361  
CGGACGGCACGGCGCGCACC GCGCGCGCGGCGCTTCGGGGCACCGGGCTCGACGGGGTGC 6420  
GCCTGCCGTGCCGCGCGTGGCGCGCGCGCGCGAAGCCCCGTGGCCCCGAGCTGCCCCACG  
. . . . .  
. . . . .  
6421  
TCAGCGGGACGTCCAACGGAAGGCAAGCCCCCGTACCCAGCCTGGTCAAGGCGCTCATCG 6480  
AGTCGCCCTGCAGGTTGCCTTCCGTTCCGGGGCATGGGTCGGACCAAGTTCCGCGAGTAGC

orfdwn3>

FIG. 1N

. . . . . M P

G

6481

CCATTCCCTGAGGAGGTCCCGCCTTGACCACAGCAATCTCCGCGCTCCCGACCGTGCCCG 6540

GGTAAGGGACTCCTCCAGGGCGGAACCTGGTGTCTGTTAGAGGCGCGAGGGCTGGCACGGGC

. . . . .

6541

GCTCCGGACTCGAAGCACTGGACCGTGCCACCCTCATCCACCCACCCCTCTCCGGAAACA 6600

S G L E A L D R A T L I H P T L S G N

T

. . . . .

6601

CCGCGGAACGGATCGTGCTGACCTCGGGGTCCGGCAGCCGGGTCCGCGACACCGACGGCC 6660

A E R I V L T S G S G S R V R D T D G

R

. . . . .

6661

GGGAGTACCTGGACGCGAGCGCCGTCCTCGGGGTGACCCAGGTGGGGCCACGGCCGGGGCCG 6720

E Y L D A S A V L G V T Q V G H G R A

E

. . . . .

6721

AGCTGGCCCCGGGTGCGGGCCGAGCAGATGGCCCCGGCTGGAGTACTTCCACACCTGGGGGA 6780

L A R V A A E Q M A R L E Y F H T W G

T

. . . . .

6781

CGATCAGCAACGACCGGGCGGTGGAGCTGGCGGCACGGCTGGTGGGGCTGAGCCCCGGAGC 6840

I S N D R A V E L A A R L V G L S P E

P

. . . . .

6841

CGCTGACCCGCGTCTACTTCACCAGCGGGGGGGCCGAGGGCAACGAGATCGCCCTGCGGA 6900

L T R V Y F T S G G A E G N E I A L R

M

. . . . .

FIG. 10

6901  
TGGCCCGGCTCTACCACCACCGGCGCGGGGAGTCCGCCCCGTACCTGGATACTCTCCCGCC 6960  
A R L Y H H R R G E S A R T W I L S R  
R

6961  
GGTCGGCCTACCACGGCGTCGGATACGGCAGCGGCGGCGTCACCGGCTTCCCCGCCTACC 7020  
S A Y H G V G Y G S G G V T G F P A Y  
H

7021  
ACCAGGGCTTCGGCCCCCTCCCTCCCGGACGTCGACTTCCTGACCCCGCCGCAGCCCTACC 7080  
Q G F G P S L P D V D F L T P P Q P Y  
R

7081  
GCCGGGAGCTGTTGCGCGGTTCCGACGTCACCGACTTCTGCCTCGCCGAACTGCGCGAGA 7140  
R E L F A G S D V T D F C L A E L R E  
T

7141 CCATCGACCGGATCGGCCCCGAGCGGATCGCGGCGATGATCGGCGAGCCGATC  
I D R I G P E R I A A M I G E P I

Sau

FIG. 1P